NOTES 20: CLT-BASED INFERENCE FOR MEANS

Stat 120 | Fall 2025

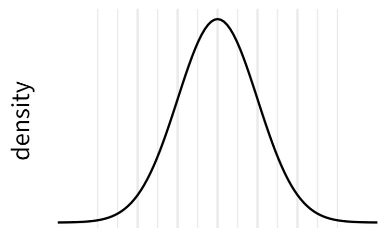
Prof Amanda Luby

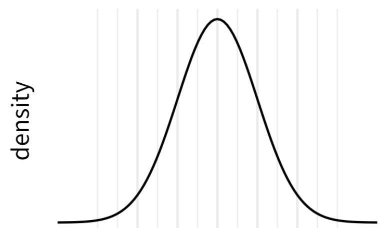
CLT: The Central Limit Theorem (CLT) tells us that if the sample size is big enough and the sample is random,

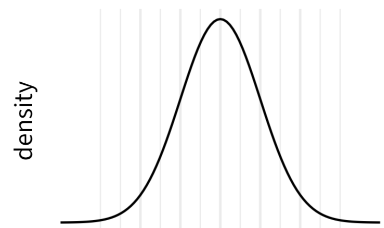
$$\bar{X} \sim N(\underline{\hspace{2em}}, \underline{\hspace{2em}})$$

Example: Finding z^* with a table. What could you say about the value of the p-value for a z-score of 2.917?

percentage	percentile (`qnorm(percentage)`)
90%	1.3
95%	1.6
97.5%	2.0
99%	2.3
99.5%	2.6







1 How big is big enough?

> **Note**
>
> **Rule of Thumb for Means**:

2 How do we find the SE?

> ℹ Note
>
> **Standard Error for Means**:
>
> **Idea:** As n gets bigger, the SE gets ________________. If ________________ is small, the SE is also ________________.

Example: Florida Lakes

H_0 :

H_A :

$\bar{X} =$

$s =$

Test stat:

p-value:

Example: Guinness Beer Acidity

H_0 :

H_A :

$\bar{X} =$

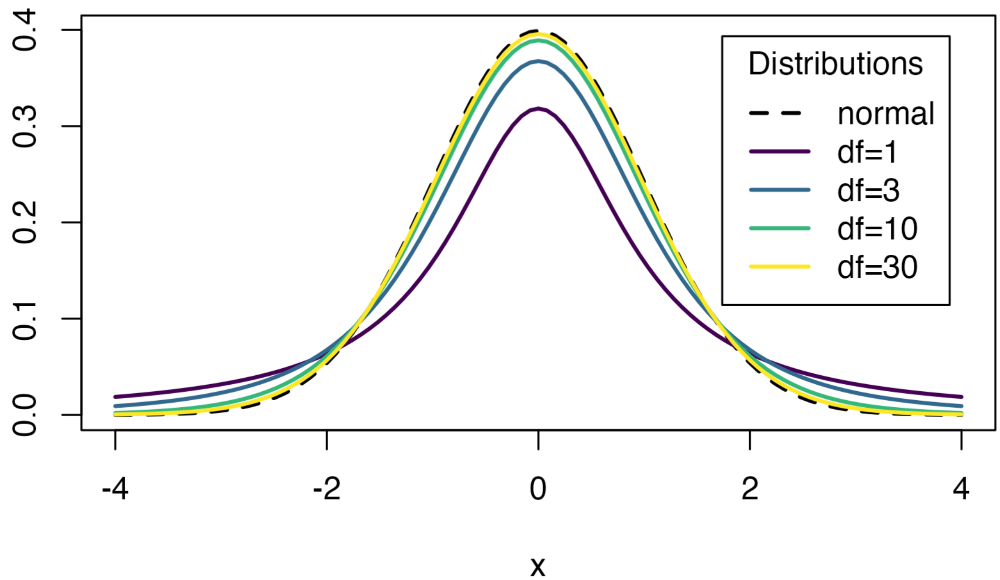
$s =$

Test stat:

p-value:

p-value (t-distribution):

3 The t-distribution



- When we divide by ____________________, the test stat has a t-distribution instead of a N(0,1)
- The t-distribution depends on the "degrees of freedom" (________________)
- When df is ________________, t-distribution has "heavier tails" than N(0,1)
- When df is ________________, the t-distribution is approximately equal to N(0,1)

Example: Florida Lakes (again)

```
t = -1.75
pt(t, df = 52)
```

```
[1] 0.04301
```

4 Summary

- Test stat for means: ___________________
- SE: ________________
- Can "safely" use the CLT if ________________
- If ________________, we can still use the CLT if there are no outliers or extreme skew
- t-distribution is better to use, but for large sample sizes it will be close to the normal distribution
- Percentage of t-distribution below t-score: `pt(t-score, df = n-1)`
- Percentile t^* for a specific percentage: `qt(percentage, df = n-1)`

5 Group Problems

1. (Adapted from Exercise 6.128)

   Plastic microparticles contaminate shorelines. Much of the pollution comes from washing fleece clothing. In a recent study, washing a fleece garment discharged on average $\bar{X} = 290$ fibers per liter of wastewater. The standard deviation was $s = 87.6$ fibers and the sample size was $n = 120$.

(a) What is the estimated *standard error* of the average number of fibers discharged per liter of wastewater when washing a fleece garment?

(b) The table below gives some percentiles of the t_{119} distribution. Use this information to construct a 99% confidence interval for the population mean. Interpret the interval in context.

percentage	percentile (`qnorm(percentage)`)
90%	1.3
95%	1.6
97.5%	2.0
99%	2.3
99.5%	2.6

(c) What sample size would we need if we wanted this interval to be *no wider* than 20?